PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68

A2

(11) International Publication Number: **WO 00/04188**

(43) International Publication Date: 27 January 2000 (27.01.00)

(21) International Application Number: PCT/US99/15836

(22) International Filing Date: 14 July 1999 (14.07.99)

(30) Priority Data:
60/092,848 15 July 1998 (15.07.98) US

(71) Applicant *(for all designated States except US)*: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608-2916 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: BERGER, Joel [US/US]; Intellectual Property – R440, P.O. Box 8097, Emeryville, CA 94622-8097 (US). RANDAZZO, Filippo [US/US]; Intellectual Property – R440, P.O. Box 8097, Emeryville, CA 94622-8097 (US). WILLIAMS, Lewis, T. [US/US]; Intellectual Property – R440, P.O. Box 8097, Emeryville, CA 94622-8097 (US).

(74) Agents: GUTH, Joseph, H. et al.; Chiron Corporation, Intellectual Property – R440, P.O. Box 8097, Emeryville, CA 94622-8097 (US).

(81) Designated States: AE, AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), EE, EE (Utility model), ES, FI, FI (Utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: METHODS AND COMPOSITIONS FOR COMPARING AND NORMALIZING ASSAYS

(57) Abstract

The invention is directed to methods and compositions for comparing results from different assays or other measurement methods and converting between their results. More specifically, the present invention is drawn polynucleotide assays and comparing and quantifying results from two different assays or assay runs. The method includes using a standard comprising a plurality of polynucleotides.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

METHODS AND COMPOSITIONS FOR COMPARING AND NORMALIZING ASSAYS

FIELD OF THE INVENTION

The invention is drawn to methods and compositions for comparing various assays or other measurement methods and normalizing their results. More specifically, the present invention is drawn to polynucleotide assays and comparing and quantifying results from two or more different assays or assay runs.

BACKGROUND OF THE INVENTION

A number of assays can measure activity or quantity using a sample that is mixed with a standard to determine the amount of binding or activation of a target molecule. These assays can utilize antibodies, receptors, proteins, or polynucleotides as targets that can bind to a number of ligands including antigens, polypeptides, organic molecules, or polynucleotides. Binding to the targets can include covalent and non-covalent association or interaction between a ligand and target to form a complex. For polynucleotide assays, the target and ligand are permitted to hybridize to form a polynucleotide duplex.

Assays using a target polynucleotide to hybridize with a sample to detect a specific sequence come in many formats. With a number of assays, the sample is bound to a solid support and the target nucleotide is labeled; examples include Southerns and Northerns.

In contrast, Dot Blots, Arrays, and branched DNA (bDNA) assays utilize a target polynucleotide bound to a solid support and the sample is labeled. Branched DNA can be used in a sandwich nucleotide hybridization assay, where the target polynucleotide is bound to a 96 well plate, for example, to capture the desired polynucleotide sequence from a sample. Next, a labeled polynucleotide probe is hybridized to the captured polynucleotide sequence to detect hybridization. Examples of a nucleotide hybridization assay are described in Urdea *et al.*, PCT WO92/02526 and Urdea *et al.*, U.S. Patent No. 5,124,246.

The convenience of Array assays is that a number of target polynucleotides can be bound to a single solid support, such as a chip. On the chip, each target polynucleotide is localized to a "patch," which creates a two dimensional matrix or array. Because each patch corresponds to a particular target polynucleotide, the locations, where polynucleotides from a sample hybridize to the chip, indicate which sequences are found in the sample. Techniques for constructing arrays and methods of using these arrays are described in EP No. 0 799 897; PCT No. WO 97/29212; PCT No. WO 97/27317; EP No. 0 785 280; PCT No. WO 97/02357; U.S. Patent No. 5,593,839; U.S. Patent No. 5,578,832; EP No. 0 728 520; U.S. Patent No. 5,599,695; EP No. 0 721 016; U.S. Patent No. 5,556,752; PCT No. WO 95/22058; U.S. Patent No. 5,631,734; Fodor *et al.*, WO 92/10588; Shalon *et al.*, WO 95/35505; Pinkel *et al.*, WO 96/17958; Augenlicht, U.S. Patent No. 4,981,783; Shalon *et al.* (1996) *Genome Research 6*:639-645; Schena *et al.* (1995) *Science 270*:467-470; Schena *et al.* (1996) *Proc. Natl. Acad. Sci. USA 93*:10614-10619.

Polynucleotide assays also are described in Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual, 2nd ed.* (Cold Spring Harbor Press, Cold Spring Harbor, NY.

SUMMARY OF THE INVENTION

An embodiment of the invention is a method of comparing and/or quantifying assay results utilizing a reference standard mixture (RSM) in an assay to detect, for example, polynucleotides. The method comprises:

(a) providing a reference standard mixture (RSM);

(b) providing a first and second sample;

(c) providing a target polynucleotide;

(d) contacting the RSM and the first sample with the target polynucleotide under conditions that permit formation of RSM-target duplexes and sample-target duplexes by hybridization;

(e) determining the presence of duplexes;

(f) determining the ratio of the amount of sample-target duplexes versus RSM-target duplexes formed;

(g) contacting the RSM and the second sample with the target polynucleotide under conditions that permit formation of RSM-target duplexes and sample-target duplexes;

(h) determining the ratio of the amount of sample-target duplexes formed versus RSM-target duplexes formed; and

(i) comparing the ratios of steps (f) and (h).

DETAILED DESCRIPTION

The instant invention relates to a RSM that is useful in a polynucleotide assay to quantify the amount of a particular polynucleotide sequence in a sample. A RSM is a single polynucleotide sequence or plurality of sequences each at a specific concentration. When combined with a sample, the RSM can be used to determine the ratio of hybridization to a target polynucleotide by polynucleotides in the sample as compared to polynucleotides in the RSM.

More specifically, the RSM of the invention comprises a plurality of polynucleotides synthesized by *In vitro* transcription, wherein the polynucleotides of the RSM are capable of hybridizing to the target polynucleotides of an assay.

The use of an RSM can overcome assay variability that is either unwanted or uncontrollable. The ability to control the composition of a RSM can overcome other assay variations when the sample and RSM are hybridized to target polynucleotides simultaneously. Sources of variability in polynucleotide assays, for example, include: conformational accessibility of target polynucleotide for hybridization; mixing conditions when contacting the sample and target polynucleotides; binding of the target polynucleotide to a solid support, if one is used.

When a sample competes for or co-hybridizes to a target polynucleotide with a RSM, the ratio of hybridization between a sample and RSM remains the same even if other factors vary in different assay runs. The ratio remains the

same because the sample and RSM are exposed to the same conditions during competition or co-hybridization for the target polynucleotide. The ratio of sample hybridization versus RSM hybridization can be used to compare different samples, if the same amount of the same RSM is used to compete or co-hybridize with different samples. Alternatively, different RSMs can be used with different samples by determining the ratio of polynucleotide concentrations between the different RSMs.

Composition of the Standard

A RSM can comprise a single component or plurality of compounds that interact with a target molecule. Polynucleotide RSMs, for example, can include a number of polynucleotides with different sequences. Polynucleotide RSMs can include, without limitation, RNA or DNA molecules, such as mRNA, cDNA, or polynucleotides comprising nucleotide analogs such as peptide nucleic acids or non-natural nucleotides such as inosine.

Typically, a polynucleotide of the present invention will comprise two nucleotides covalently linked. Generally, the linkages are phosphodiester bonds, although in some cases, nucleic acid analogs are included that may have alternative backbones, comprising, for example, phosphoramide Beaucage *et al.* (1993) *Tetrahedron 49(10)*:1925 and references therein; Letsinger (1970) *J. Org. Chem. 35*:3800; Sprinzl *et al.* (1977) *Eur. J. Biochem. 81*:579; Letsinger *et al.* (1986) *Nucl. Acids Res. 14*:3487; Sawai *et al.* (1984) *Chem. Lett. 805*; Letsinger *et al.* (1988) *J. Am. Chem. Soc. 110*:4470; and Pauwels *et al.* (1986) *Chemica Scripta 26*:141, phosphorothioate, phosphorodithioate, O-methylphosphoroamidite linkages (see Eckstein, *Oligonucleotides and Analogues: A Practical Approach,* Oxford University Press), and peptide nucleic acid backbones and linkages (see Egholm (1992) *J. Am. Chem. Soc. 114*:1895; Meier *et al.* (1992) *Chem. Int. Ed. Engl. 31*:1008; Nielsen (1993) *Nature 365*:566; Carlsson *et al.* (1996) *Nature 380*:207.

In a RSM, the specific polynucleotide sequences and the amount of each polynucleotide sequence are pre-determined. A library of mRNA isolated from

a cell without further characterization or normalization is not considered a RSM. A mRNA library from a cell is a not a delimited population of mRNA sequences because the cell generates only a subset of possible mRNA sequences from its genome and the level of transcription can be variable. The sequences and the amount of each sequence in a mRNA library from a cell can vary depending on tissue type, cell type, the culturing conditions and length in culture. Thus, cellular mRNA may not comprise sequences that are capable of hybridizing to a target polynucleotide of an assay. Further, a composition of mRNA can still remain undefined after mixing a number of undefined cellular mRNA libraries in an attempt to "average out" any discrepancies that occur from uncontrollable variations of cell growth or culturing.

In contrast, a RSM of the instant invention comprises a population of sequences each at a specific concentration. *In vitro* transcription systems and chemical synthesis allow production of any desired polynucleotide sequence. In a preferred embodiment, the RSMs of the invention are produced by *In vitro* transcription reaction(s) using the target polynucleotides of the polynucleotide assay as templates for the transcription reaction.

The size of the polynucleotide sequences of the RSM can be chosen based on convenience. Typically, the size of the polynucleotide sequences is at least 8 bases; more typically at least 12, 16, or 18 bases; even more typically, 20, 30, or 50 bases long. Usually, the size of the polynucleotide is no more than about 7,500 bases; more usually no more than either 5,000; 4,000; or 3,500, bases; even more usually, no more than 3,000; 2,000; or 1,000 bases. Preferably the size of the polynucleotide is between about 2,500 bases and 150 bases; more preferably, between about 2,000 bases and 500 bases; even more preferably, between about 1,000 bases to about 200 bases.

The sequence of the polynucleotides in the RSM can include variability at some positions, i.e., some position may be degenerate by permitting more than one base. Typically, the reverse complement of each sequence of the RSM exhibits at least 75% sequence identity to a target polynucleotide in the assay; more typically at least 80% sequence identity; even more typically, at least 85%

sequence identity. Preferably, the reverse complement of each sequence in the RSM exhibits at least 90% sequence identity to a target polynucleotide in the assay; even more preferably, at least 95%; even more preferably, 98%; even more preferably 99%; even more preferably, 100% sequence identity.

To determine sequence identity, target sequences and polynucleotide sequences of the RSM can be aligned using the methods and computer programs, including BLAST, available over the world wide web at http://ww.ncbi.nlm.nih.gov/BLAST/. Another alignment algorithm is Fasta, available in the Genetics Computing Group (GCG) package, from Madison, Wisconsin, USA, a wholly owned subsidiary of Oxford Molecular Group, Inc. Other techniques for alignment are described in Methods in Enzymology, vol. 266: Computer Methods for Macromolecular Sequence Analysis (1996), ed. Doolittle, Academic Press, Inc., a division of Harcourt Brace & Co., San Diego, California, USA. Preferably, an alignment program that permits gaps in the sequence is utilized to align the sequences. The Smith-Waterman is one type of algorithm that permits gaps in sequence alignments. See *Meth. Mol. Biol. 70*:173-187 (1997). Also, the GAP program using the Needleman and Wunsch alignment method can be utilized to align sequences.

Preferably, the FastDB algorithm is used to determine sequence identity. FastDB is described in Current Methods in Sequence Comparison and Analysis, *Macromolecule Sequencing and Synthesis, Selected Methods and Applications*, pp. 127-149, 1988, Alan R. Liss, Inc. Percent Identity is calculated by FastDB based upon the following parameters:

Mismatch Penalty:	1.00;
Gap Penalty:	1.00;
Gap Size Penalty:	0.33; and
Joining Penalty:	30.0.

One parameter for determining percent sequence identity is the percentage of the alignment region length where the strongest alignment is found.

The percentage of the alignment region length is calculated by counting the number of residues of the individual sequence found in the region of strongest alignment. This number is divided by the total residue length of the target or RSM polynucleotide sequence to find a percentage. An example is shown below:

```
Target sequence:    GCGCTAAATACTCACTCGAGG
                        |   ||| |||| |||
RSM sequence:       TATAGCCCTAC.CACTAGAGTCC
                    1   5   10    15
```

The region of alignment begins at residue 9 and ends at amino acid 19. The total length of the target sequence is 20 residues. The percent of the alignment region length is 11 divided by 20 or 55%, for example.

Percent sequence identity is calculated by counting the number of residue matches between the target and RSM polynucleotide sequence and dividing total number of matches by the number of residues of the target or RSM sequence found in the region of strongest alignment. For the example above, the percent identity would be 10 matches divided by 11 residues, or approximately, 90.9%

The percent of the alignment region length, typically, is at least about 55% of total length of the sequence; more typically, at least about 58%; even more typically; at least about 60% of the total residue length of the sequence. Usually, percent length of the alignment region can be as much as about 62%; more usually, as much as about 64%; even more usually, as much as about 66%.

Alternatively, sequence similarity can be assessed empirically. For example, polynucleotide sequence of a RSM will hybridize to a target polynucleotide sequence of the assay under stringent conditions to form stable duplexes between homologous regions. Stable duplexes are those, for example, which would withstand digestion with a single-stranded specific nuclease(s), such as S_1. Such duplexes can be analyzed by various methods, such as size determination of digested fragments.

"Hybridization" refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Two sequences will be placed in contact

with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Vol. 2, chapter 9, pp. 9.47-9.57.

"Stringency" refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 12^0 to 20^0 C below the calculated T_m of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook, *et al.*, above at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the target and the sequences being detected. The total amount of the polynucleotides to be studied can vary a magnitude of 10, from 0.1 to 1 μg for a plasmid or phage digest to 10^{-9} to 10^{-8} μg for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of a target polynucleotide can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 μg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a target polynucleotide radiolabeled with 10^8 cpm/μg. For a single-copy mammalian gene a conservative approach would start with 10 μg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a target

polynucleotide radiolabeled with greater than 10^8 cpm/μg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature (T_m) of a DNA-DNA hybrid between the target and sequence of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the target is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

Tm= 81 + 16.6(log10C_i) + 0.4[%G + C)]-0.6(%formamide) - 600/n-1.5 (% mismatch). Where C_i is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth and Wahl (1984) *Anal. Biochem. 138*:267-284).

In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (i.e., stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the labeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a target polynucleotide with 95% to 100% sequence identity to the sequence to be detected, 37°C for 90% to 95% sequence identity, and 32°C for 85% to 90% sequence identity. For lower percentage sequence identity, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the

target polynucleotide and the sequence to be detected are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If nonspecific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

Stringent conditions include hybridization in a solution of at least about 5 x SSC at 65°C, or at least about 4 x SSC at 42°C; see, for example, U.S. Patent No. 5,707,829.

Sequences to Decrease Non-Specific Binding

The RSM can comprise polynucleotide sequences that bind to highly repetitive elements that exist in naturally occurring genomes, such as alu, poly A^+ or Human Cot-1 sequences. These sequences can be blocked using complementary sequences.

Preferably, a poly A^+ to decrease non-specific binding comprises at least approximately 150 adenine nucleotides, more preferably, at least 100 nucleotides; even more preferably, 80 nucleotides.

Other oligonucleotides can be added that comprise sequence that may common to the sample polynucleotides. For example, if all the target polynucleotides comprise vector sequence, then an oligonucleotide that binds to the common vector sequences may be useful to prevent non-specific binding.

Concentration of the Polynucleotide Sequences

The concentration of polynucleotides possessing a desired sequence is defined in the RSM of the instant invention. The concentration of the polynucleotides can be controlled by adding a specific concentration of each polynucleotide sequence to one mixture. Alternatively, a plurality of template polynucleotides can be added at a pre-determined concentration to an *In vitro* transcription system, and transcription of each polynucleotide can occur simultaneously. The polynucleotide concentrations resulting from the *In vitro*

transcription can be later adjusted to meet pre-determined concentrations or to normalize against a reference.

The concentration of each polynucleotide sequence is generally within the standard error or background of the assay format used. For example, if the assay format can only determine the concentration of polynucleotides within a range of ±10%, then the concentration of each polynucleotide sequence in the RSM is within ±10% of the desired concentration. Typically, for an Array format, the concentration of each polynucleotide sequence in the RSM is at least 2 fold of the desired concentration; more typically within ±100% of the desired concentration; even more typically, within ±50%; even more typically, within ±25%; even more typically, within ±15%; even more typically, within ±10% of the desired concentration.

Alternatively, the range of the desired concentration can be determined by the background of the assay. For example, if the negative control gives rise to 100,000 fluorescent volumes in an Array format assay, then the desired concentration of each polynucleotide sequence produces a signal that is more than 100,000 fluorescent units.

Another factor to consider is the variability of the assay. For example, the fluorescent signal measured from a sample in an Array format assay may vary from sample to sample although each sample has the same amount of DNA. Thus, the concentration of the polynucleotide sequence in the RSM may be chosen to produce a particular range of signal intensities rather than a single intensity. Usually, the range of signal intensities can vary from 2 to 10 fold; more usually, 2 to 8 fold; even more usually, 2 to 5 fold.

One factor that can be used to determine the desired concentration of each sequence in the RSM of polynucleotides is the linear response of the assay or detection of the instruments used. Typically, the concentration chosen produces a signal that is above the background and below the saturation or high end of the curve. For example, each polynucleotide sequence can be at a concentration that can be detected within the middle of the linear range of the assay.

The linear range of an assay is the portion of the standard curve where there is a one-to-one correlation between the concentration and the signal that is measured, such as a fluorescent signal. In this linear range, the measurement of signal corresponds to only one concentration. This correspondence can be approximated by any type of curve, such as linear, polynomial, logarithmic, or sinusoidal, for example.

Typically, each sequence produces a signal within the middle of the linear range, more typically, within the middle two quadrilles of the range when the linear range is divided into four equal portions, even more typically, within the second quadrille of the range. For example, if the standard curve is linear within 100,000 to 1,000,000 fluorescent volumes in an Array format assay, then each polynucleotide sequence of a RSM is usually at a concentration that produces a signal between about 325,000 and about 775,000 fluorescent volumes; more usually, each polynucleotide sequence is at a concentration that produces a signal of between about 325,000 and about 550,000 fluorescent volumes; even more usually, about 500,000 fluorescent volumes.

Alternatively, the concentration of the desired polynucleotides can be determined by the linear response of the instrument that detects signal of the labeled polynucleotides. The linear response of an instrument is the region in which the signal measured correlates with only one particular polynucleotide concentration. For example, a fluorescent scanner can measure up to 50,000,000 fluorescence volumes before polynucleotide sample or RSM as a whole produces a saturating signal. Thus, the linear range of detection is within 0 to 50,000,000 fluorescent volumes. Using these guidelines, each polynucleotide sequence in a RSM is usually at a concentration that produces a signal between about 0 and about 25,000,000 fluorescent volumes; more usually, each polynucleotide sequence is at a concentration that produces a signal of between about 0 and about 12,500,000 fluorescent volumes; even more usually, about 500,000 fluorescent volumes.

In another example, the linear range of the Array format assay correlates a signal to a concentration of about 0.308 ng to about 3.08 ng of a particular

sequence for every microgram of total polynucleotide sample. Given this linear range, each polynucleotide sequence in the RSM is preferably at a concentration of about 1.0 ng to about 2.39 ng for every microgram of RSM; more preferably, about 1.0 ng to about 1.69 ng; even more preferably, about 1.5 ng. Preferably, each nucleotide sequence is at a concentration of about 1.0 ng to about 2.0 ng for every microgram of a RSM. If the polynucleotide sequence is greater than 1.0 kilobases (kb), then the concentration of the polynucleotide sequence can be higher than 1.0 ng to 2.0 ng. If the polynucleotide sequence is less than 1.0 kb, the concentration of the polynucleotide sequence can be lower than 1.0 ng to 2.0 ng.

In another example, the linear range of the curve correlates a signal to about 50 to about 500 copies of a transcript in the total cellular polynucleotide population (also referred to as copies per cell). Given this linear range, each polynucleotide sequence in a RSM usually is at a concentration of about 162.5 to about 387.5 copies per cell; more usually, at a concentration of about 162.5 to about 275 copies per cell. The population of all cellular polynucleotides can refer to either all genomic DNA, all mRNA, or all polyA$^+$ mRNA. Copies per cell can be converted to micrograms assuming 10^7 cells produce approximately 1 μg of polyA$^+$ RNA and each transcript is approximately 1000 bases.

Usually, the concentration of the polynucleotide sequence in the RSM is not more than 5 fold lower than the concentration of the polynucleotide in the sample; more usually, not more than 1 to 3 fold lower than the concentration of the polynucleotide in the sample.

Determining the ratio of hybridization between a sample and two or more different RSMs can be useful. For example, the sample can be compared to either a "high" or "low" concentration RSM hybridization to a target polynucleotide in separate hybridization experiments. If the polynucleotide concentration in the sample is more than 3 to 5 fold lower than the sequence in the "high" standard, the sample will be within 3 to 5 fold of the "low" standard. Alternatively, the sample can be diluted until the concentration of the sequence

to be assayed is within 1 to 3 fold of the concentration of the sequence in the RSM.

A "high" RSM can comprise polynucleotide sequences at a concentration that produces a signal in the top quadrille of the linear range of the standard curve. A "low" RSM can comprise polynucleotide sequences at a concentration that produces a signal in the first or second quadrille of the linear range of the standard curve, more preferably in the second quadrille.

Also, a sample can be compared to two different RSMs that bind the different portions of the same target polynucleotides. For example, a "5' RSM comprises polynucleotides that are complementary to the 5' half the target polynucleotides of the assay. The "3'" RSM comprises to polynucleotides that are complementary to the 3' half the target polynucleotides of the assay.

Synthesis of Polynucleotides

The polynucleotides for the RSM can be constructed by a number of techniques including chemical synthesis or *In vitro* transcription methods to produce RNA or DNA or polynucleotide analogs.

Polynucleotides, including those with non-natural bases, with sugar phosphate backbones can be synthesized with the phosphoramidite method of oligonucleotide synthesis, for example. See Beaucage *et al.* (1981) *Tet. Lett.* *22*:1859-1862 and U.S. Patent No. 4,668,777. Automated devices for synthesis are available to create oligonucleotides using this chemistry. Examples of such devices include Biosearch 8600, Models 392 and 394 by Applied Biosystems, a division of Perkin-Elmer Corp., Foster City, California, USA; and Expedite by Perceptive Biosystems, Framingham, Massachusetts, USA. Synthetic RNA, phosphate analog oligonucleotides, and chemically derivatized oligonucleotides can also be produced, and can be covalently attached to other molecules.

RNA oligonucleotides can be synthesized, for example, using RNA phosphoramidites. This method can be performed on an automated synthesizer, such as Applied Biosystems, Models 392 and 394, Foster City, California, USA.

See Applied Biosystems User Bulletin 53 and Ogilvie *et al.* (1987) *Pure & Applied Chem. 59*:325-330.

Phosphorothioate oligonucleotides can also be synthesized. A sulfurizing reagent, such as tetraethylthiruam disulfide (TETD) in acetonitrile can be used to convert the internucleotide cyanoethyl phosphite to the phosphorothioate triester within 15 minutes at room temperature. TETD replaces the iodine reagent, while all other reagents used for standard phosphoramidite chemistry remain the same. Such a synthesis method can be automated using Models 392 and 394 by Applied Biosystems, for example.

Oligonucleotides of up to 200 nucleotides can be synthesized, more typically, 100 nucleotides, more typically 50 nucleotides; even more typically 30 to 40 nucleotides. These synthetic fragments can be annealed and ligated together to construct larger fragments. See, for example, Sambrook *et al.*, *supra.*

Peptide nucleic acids (PNA) comprise monomer linked by an amide bond, such as those bonds that link peptides in proteins. Such PNA molecules can be constructed using conventional peptide chemistry and instruments. Such chemistry and reagents are available from Novabiochem, San Diego, California, USA.

In vitro methods of constructing polynucleotides include Polymerase Chain Reaction (PCR), *In vitro* transcription, and nick translation.

PCR is described in Mullis *et al.* (1987) *Meth. Enzymol. 155*:335-350; U.S. Patent No. 4,683,195; and U.S. Patent No. 4,683,202. Two primer polynucleotides hybridize with the template nucleic acids and are used to prime the reaction. A thermostable polymerase creates copies of template nucleic acids from the primers.

Other methods for producing DNA include a single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides as described by Stemmer *et al.* (1995) *Gene (Amsterdam) 164(1)*:49-53. In this method, assembly PCR (the synthesis of long DNA sequences from large numbers of oligodeoxyribonucleotides (oligos)) is acheived. The method is

derived from DNA shuffling (Stemmer (1994) *Nature 370*:389-391), and does not rely on DNA ligase, but instead relies on DNA polymerase to build increasingly longer DNA fragments during the assembly process. For example, a 1.1-kb fragment containing the TEM-1 beta-lactamase-encoding gene (bla) can be assembled in a single reaction from a total of 56 oligos, each 40 nucleotides (nt) in length. The synthetic gene can be PCR amplified and cloned in a vector containing the tetracycline-resistance gene (Tc-R) as the sole selectable marker. Without relying on ampicillin (Ap) selection, 76% of the Tc-R colonies were Ap-R, making this approach a general method for the rapid and cost-effective synthesis of any gene.

RNA or DNA can be produced by *In vitro* transcription from a template polynucleotide using commercial reagents and kits available from New England Biolabs, Beverly, Massachusetts, USA; Invitrogen Corporation, San Diego, California, USA; and Ambion, Incorporated, Austin, Texas, USA, for example. To utilize *In vitro* transcription reactions, the desired template is constructed by operably linking a target polynucleotide sequence to a promoter that is recognized by polymerase to produce either DNA or RNA. Examples of promoters include: the T3 phage promoter; the T7 phage promoter; and the SP6 phage promoter. If the Ambion, Inc. MEGAscript™ T7 kit, catalog no. 1334, is used, the polynucleotide sequence is operably linked to a T7 phage promoter. Ambion, Inc. is located in Austin, Texas, USA.

Labeling of polynucleotides

Typically, the polynucleotides of the RSM are labeled to permit detection of its hybridization to the target polynucleotide. Such labels can include, without limitation, radioactive isotopes; fluorescent labels, and polypeptides detectable by antibodies; biotin detectable by a labeled avidin conjugate; chemiluminescent label; enzymes; substrates; cofactors; inhibitors; magnetic particles; heavy metal atoms; and spectroscopic labels. Patents teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241.

Useful labels in the present invention include fluorescent dyes, such as fluorescein isothiocyanate, texas red, rhodamine, and the like.

The labels can be incorporated after the polynucleotide sequence is synthesized or labeled nucleotide analogs can be incorporated during the synthesis of the polynucleotide sequence. Alternatively, the sample can be transcribed with a primer that contains a sequence that is capable of hybridizing to an "amplifer," which includes a multiple labels.

Fluorescent nucleotides analogs, such as a fluorsescein-12-deoxycytosine triphosphate (dCTP) or lissamine-5-dCTP are commercially available, from New England Nuclear, (DuPont), Boston, Massachusetts, USA can be used to label polynucleotides. For the Array assay format, preferred fluorescent labeled nucleotide analogs include Cy2, Cy3, Cy3.5, Cy5, and Cy5.5, available from Amersham Pharmacia Biotech, Uppsala, Sweden. Most preferably, Cy3 and Cy5 are used. Preferably, Cy3 and Cy5 conjugated to dCTP is used.

The polynucleotide sample need not be labeled when the sample and RSM are competing for a limited amount of target polynucleotide. However, when the target polynucleotide is in excess, the polynucleotide sample is preferably labeled with a marker that is distinguishable from the label used with the RSM.

Different labels can produce different signal intensities. Normalization of the signal intensities may be desired when computing ratios. For example, the fluorescent intensities are not the same for a sample comprising equal amounts of Cy3 or Cy5-labled DNA. Cy5 produces a less intense signal than Cy3. The Cy5 signal can be multiplied by a normalization factor to produce a 1:1 ratio between Cy3 and Cy5 fluorescent signals when the concentration of the Cy3-labeled DNA is the same as the concentration of the Cy5-labeled DNA. A Cy5-labeled polynucleotide can be 1/10 to ½ as intense as a Cy3-labeled polynucleotide.

The normalization factor can be determined empirically as shown in Example 2. In one Array assay format the normalization factor is approximately 3.33. The factor can vary from 1 to 10.

Assay Formats

Assay formats for detecting and quantifying polynucleotides can be either liquid or solid phase formats. Assay formats include branched DNA, Northerns, Southerns, Dot Blots, and Array formats. For these formats, the target polynucleotide can be bound to solid supports that include, for example, plastic, nitrocellulose, nylon, or glass slides. Also, activated substrates or other surfaces that provide attachment sites for target polynucleotides can be utilized. For example, gold slides can be used that permit monolayer formation of thiol-alkanes that can incorporate sites to which target polynucleotides can be attached.

Further, either a single polynucleotide sequence or a number of polynucleotide sequences can be assayed alone or simultaneously. In any of these formats, the RSM of the instant invention is a useful means for comparing results from different assay runs.

A RSM with a plurality of polynucleotide sequences is of particular use when a sample is being assayed simultaneously for a number of sequences in an Array format. The Array format can take the form of a Dot Blot, where a number of target polynucleotide sequences are bound to a nitrocellulose substrate.

For high throughput, a high density array format can be utilized. "High density" Array format is an array of regions having a density of discrete regions of at least about 100/cm^2, and preferably at least about 1000/cm^2. The regions in a microarray have typical dimension, e.g., diameters or widths, in the range of between about 10-250 μm, and are separated from other regions in the array by about the same distance.

The target polynucleotide can be excess or can be limiting in these polynucleotide formats. In either case, the RSM is a useful standard to use in a

ratio with any sample to compare results between assay runs or between different samples.

With the availability of a number of different labels, more than one sample could be mixed with more than one RSM to hybridize to a target polynucleotide(s). For example, a single sample could be co-hybridized with a "high" and "low" concentration RSM. Alternatively, a single sample could be diluted and different dilution could be incorporate different labels; and then all the dilution could be mixed together with a single RSM before hybridizing to a target polynucleotide(s). In another embodiment, two or more samples with different labels could be mixed with a single RSM and then hybridized to target polynucleotide(s).

EXAMPLES

Example 1

Synthesizing a RSM

RNA was produced from template DNA synthesized from clones available from the I.M.A.G.E. Consortium. RNA produced from DNA templates, which were aliquoted into a mixture at a fixed concentration of 1-2 ng per microgram of total DNA were used for the RNA transcription. Either this RNA mixture or single stranded DNA transcribed from the RNA can be used as a RSM.

Producing Template DNA Comprising a T7 Promoter

The following clones were used to produce a DNA template for a transcription synthesis to produce RNA:

Image ID	Gene Name	Plate Location
22012	Human ADP-ribosylation factor (ARF3) mRNA	A2
22493	H. sapiens mRNA for ribosomal protein L26	A5
28410	Human AMP deaminase (AMPD2) mRNA	B6
28985	Human alpha-N-acetylgalactosaminidase mRNA	B7
30272	Human Na/H antiporter (APNH1) mRNA	B9
37451	Homo sapiens creatine kinase B mRNA	D4
39173	H. sapiens mRNA for a cell surface protein	D9
39285	Human liver glutamate dehydrogenase mRNA	D10
40026	Human mitochondrial ADP/ADT translocator mRNA	E2
40304	Human mRNA for flavoprotein subunit of complex II	E3
40360	Human eukaryotic initiation factor 2B-epsilon mRNA	E4
41199	Human chromatin assembly factor-1 p60 subunit mRNA	E5
42214	Homo sapiens protein tyrosine kinase (Syk) mRNA	E10
43021	Human mRNA for histidyl-tRNA synthetase (HRS)	F2
43129	Human topoisomerase I mRNA	F3
43622	Human glutamate receptor 2 (HBGR2) mRNA	F9
45941	H. sapiens mRNA for DNA (cytosin-5)-methyltransferase	G3
50359	H. sapiens PMI1 mRNA for phosphomannose isomerase	G10
50941	Homo sapiens cadherin-13 mRNA	G12
51666	Human DNA repair helicase (ERCC3) mRNA	H2

These clones are available from American Type Culture Collection (ATCC), an authorized I.M.A.G.E. Consortium distributor. The ATCC is located at 10801 University Boulevard, Manassas, Virginia, USA, 20110-2209.

Template DNA from these clones was produced by PCR using the following 5' primer to incorporate the T7 promoter:

TAATACGACTCACTATAGGGGAAACAGCTATGACCATGATTA CGC. (The sequence is from 5' to 3'.

The 3' primer sequence is as follows from 5' to 3':

ACGGCCAGTGAATTCCCCTT.

For the PCR, about 1ng of DNA was used with 30 picomoles of each primer. The first PCR cycle was 97°C for fifteen seconds, 55°C for 30 seconds, 72°C for two minutes. The next 29 cycles were: 97°C for five seconds, 55° for 30 seconds, 72°C for two minutes. The reaction was completed with an incubation at 72°C for five minutes. The reaction was stored at 4°C.

PCR products were purified using a kit from Qiagen, Hilden, Germany.

Transcribing RNA from the DNA Templates

RNA was transcribed from the DNA templates described above using the MEGAscript™ *In vitro* transcription kit, catalog no. 1334 available from Ambion, Inc., Austin, Texas, USA.

RNA was produced from a reaction mix as follows:

Amount	Component
2 µl	10X Reaction Buffer
2 µl	ATP Solution (75 mM)
2 µl	CTP Solution (75 mM)
2 µl	GTP Solution (75 mM)
2 µl	UTP Solution (75 mM)
6 µl	~ 1 µg of linearized template DNA
2 µl	Enzyme Mix
20 µl	q.s. with Nuclease-free water

The reaction mixture was incubated at 37°C for 4 hours, the first hour in a water bath and the last three hours in an incubator. The mixture was stored at -20°C.

The reaction was stopped by adding 115 µl of nuclease-free water and 15 µl of Ammonium Acetate Stop Solution and mixed thoroughly. The mRNA was extracted once with an equal volume of water- or buffer-saturated phenol/chloroform, and once with an equal volume of chloroform.

The RNA was precipitated by adding one volume of isopropyl alcohol and mixing well. The mixture was chilled for at least fifteen minutes at -20°C. The mixture was centrifuged at 4°C for 25 minutes at maximum speed to pellet the RNA. The supernatant was carefully removed and the precipitate was washed once with 100µl of 70% ethanol. The pellet was centrifuged again for five minutes. The supernatant was removed and the pellet was air dried. The pellet was resuspended in 35 µl of 0.1 mM EDTA RNase free.

Isolating mRNA from a Cell

RNA can be isolated from a cell to be used as sample using the ToTally RNA™, Total RNA Isolation Kit, catalog no. #1910, available from Ambion, Austin, Texas, USA.

Producing Single Stranded DNA from the mRNA Sample or RNA RSM

A mRNA sample or RSM was transcribed into single stranded DNA, to be used as in an Array format assay. Transcription is performed as described below, incorporating a fluorescently labeled nucleotide analog.

One microgram of purified poly A^+ RNA from a sample or RSM was used for the reverse transcription reaction. The RNA was thawed on ice. Next, 1 μg RNA was added to a tube with 1 μl of anchored $dT_{(25)}$ primer (20x) at a concentration of 8 μM as provided in kit no. RPK 0140 (NIF 1265) from Amersham Pharmacia Biotech, Uppsala, Sweden. Also, sufficient RNAse free water was added to the tube to take the reaction to a final volume of 10.5 μl when the components in Table 1 below have been taken into account.

The primer was pre-annealed to the RNA by heating the solution in the tube to 70°C for five minutes and then cooled to room temperature over ten minutes. Next, the components in Table 1 were added to the microfuge tube.

Table 1

Volume	Component
4 µl	SuperScript II 5x First Strand Buffer: 250mM Tris-HCl, pH 8.3, 375 mM KCl, and 15 mM $MgCl_2$ Available as catalog no. 18064-014 from Life Technologies, a producer of Gibco-BRL products and located in Rockville, Maryland, USA.
2 µl	0.1 M dithiothreitol (DTT)
1 µl	dNTP mix of: 2 mM dATP, 2 mM dGTP, 2 mM dTTP, and 1 mM dCTP.
1 µl	1 mM Cy3-dCTP or Cy5-dCTP These nucleotide analogs are available from Amersham Pharmacia Biotech, Uppsala, Sweden.
0.5 µl	RNaisin at 20-40 units/µl Available from Promega Corporation, Madison, Wisconsin, USA.

DNA transcribed from the RSM was labeled with Cy5 and the DNA from the mRNA sample was labeled with Cy3. The solution in the microfuge tube was mixed gently and spun for a few seconds in a microcentrifuge.

One microliter of SuperScript II reverse transcriptase at a concentration of 200 units/µl or 600 units/µl was added, catalog no. 18064-014 or 18064-105, respectively, available from Life Technologies, a producer of Gibco-BRL products and located in Rockville, Maryland, USA.

The solution was mixed gently and spun again. The solution was incubated at 42°C for 2-2.5 hours. The solution was centrifuged briefly and placed on ice before RNA degradation and purification or is stored at -20°C immediately after the incubation was complete.

RNA was removed from the reverse transcription products by adding the following:

10μl of 1 μL of RNAse H at 1 unit/μl diluted in 9 μl of 1X First Strand Buffer; and 1 μl RNAse One at 1 unit/μl in 1X Reaction Buffer. The RNAse H is available from Boehringer Mannheim, Indianapolis, Indiana, USA. RNAse One is available from Promega Corporation, Madison, Wisconsin, USA.

Alternatively, the RNA can be removed by alkaline hydrolysis of the RNA template by heating the solution at 94°C for three minutes. Next, 1 μl of 5 M sodium hydroxide was added and the solution was incubated at 37°C for ten minutes. One microliter of 5 M hydrochloric acid and 5 μl 1 M Tris-HCl, pH 6.8 was added and the solution was mixed and centrifuged briefly.

Purification of the Single Stranded DNA Reverse Transcribed from RNA

DNA reversed transcribed from 1 μg of the RSM and DNA transcribed from 1 μg of the RNA sample was mixed together and then further purified over a GFX column, catalog no. 27-9602-01, available from Amersham Pharmacia Biotech, Uppsala, Sweden. The DNA was loaded onto the column as per manufacturer's instructions. The DNA was eluted with 30-50 μl of TE (5 mM Tris, 0.05 mM EDTA) pH 8.0 or autoclaved double-distilled water adjusted to pH 8.4. The eluted material was dried down in a speed vacuum to a volume of 0-5 μl and resuspend to a volume of 5μl in TE or double-distilled water.

Preparing a Slide Comprising Target DNA

Ten microliters of purified PCR DNA was added to 10 μl dimethlysulfoxide (DMSO). This mixture was spotted onto coated glass slides available as MicroArray Slides Type IV, catalog no. RPK 0163, from Amersham Pharmacia Biotech, Uppsala, Sweden. The DNA was spotted down with a Array Spotter, Generation II, available from a Molecular Dynamics and Amersham Pharmacia Biotech Venture, Sunnyvale, California, USA.

The target DNAs were air dried at 20°C and at 40-60% humidity for approximately 30-60 minutes. Next, the target DNAs were crosslinked to the slide by placing the spotted slide in a UV Stratalinker 1800 available from

Stratagene, La Jolla, California, USA. The Stratalinker was set on auto-cross-link (1200 μJoules x 100).

Next, the slide was baked at 80°C for one hour. The slides were stored under nitrogen.

Hybridizing RSM and Sample to a Target Polynucleotide in an Array Format

In brief, the slide comprising the target polynucleotides is exposed to pre-hybridization solution before hybridization with the RSM and sample polynucleotides. Next, the sample and RSM were taken up in a small volume of hybridization solution and pre-annealed to DNA to block non-specific binding. Following heat denaturation, the hybridization mix is deposited on the slide surface and then a cover slip is dropped over the mix. Hybridization is performed overnight in a humid chamber. The following day, the slide is washed briefly at low stringency and more extensively at higher stringency. The slide is rinsed briefly in distilled water to remove traces of salt and SDS (which if dried on the slide, result in background fluorescence) and then immediately dried with ultra high purity nitrogen gas.

More specifically, the slide comprising the target DNA was incubated in 5x SSC, 0.1% SDS, and 50% deionized formamide at 42°C for 3-4 hours. The slide was then washed briefly with water twice and dried with compressed air.

Next, 5 μl solution of the sample and RSM were added to the following:

Volume	Component
0.5 μl	Primer Solution of: 459 pmoles/μl of oligonucleotide with the sequence: GAAACAGCTATGACCATGATTACGCC in TE.
1 μl	Poly A DNA comprising 80 nucleotides (approximately 1.25-1.4 μg)
1 μl	Cot 1 Human DNA (approximately 1 μg) Available as catalog no. 1 581 074 from Boehringer Mannheim, Indianapolis, Indiana, USA.

This solution was incubated at 95°C for one minute and then was placed on ice. The remaining components were then added to a final concentration of

5x SSC, 0.1% SDS, and 50% deionized formamide (checked pH of formamide that it was not acidic) to a final volume of thirty microliters. This thirty microliter solution was incubated at 42°C for thirty minutes.

The solution was pipetted onto the slide and the cover slip dropped down onto it. (Cover slips were wiped with 70% ethanol and dried.) Next, the slide was placed in a humid chamber to permit hybridization overnight at 42°C with 1x Hybridization Solution at the bottom of the chamber.

The cover slip was removed, and the slide was washed in Wash Solution 1 for five minutes at room temperature. Wash Solution 1 was poured off, and the slide was washed twice in Wash Solution 2 for ten minutes at room temperature. The slide was rinsed in distilled water once or twice and was immediately dried with ultra high purity nitrogen gas.

Wash Solution 1 was 1X SSC and 0.2% SDS. Wash Solution 2 was 0.1X SSC and 0.2% SDS.

Measuring Fluorescence

The dried slide was scanned in an Array Scanner, Generation II, available from available from a Molecular Dynamics and Amersham Pharmacia Biotech venture, Sunnyvale, California, USA.

The slide is scanned twice; once, to measure the red fluorescent signal produced by the Cy5 labeled RSM polynucleotides and once, to measure the green fluorescent signal produced by the Cy3 labeled sample. The red signal is measured at a PMT voltage of 750 when Cy5 is excited at a wavelength of 633 nm and measured with Filter 2 to measure the maximum fluorescent emissions at 670 nm. The green signal is measured at a PMT voltage of 700 when Cy3 is excited at a wavelength of 532 nm and measured with Filter 1 to measure the maximum fluorescent emissions at 570 nm.

Sample and RSM Preparation

An RSM was synthesized using DNA generated from each of the twenty I.M.A.G.E. clones described in Example 1 as well as Arabidopsis cDNA as a

positive control. The I.M.AG.E clones also referred to as the NIH20. This RSM was used to assay samples using each of the NIH20 clones as target polynucleotides bound to a glass slide in an Array format.

This RSM was mixed with two samples, named SM+NIH20 and SM, to determine each ratio between polynucleotide concentration of the sample and RSM.

The ratio of polynucleotide concentrations SM+NIH20 compared to SM was determined both by comparing the ratios obtained using the RSM. The ratio of SM+NIH20 to SM also was determined empirically by co-hybridizing SM+NIH20 and SM with each of the NIH20 clones as target polynucleotides.

To produce the RSM and the two samples, each of the following RNA components were synthesized:

Component Name	Method of Synthesis
Arab.	2.9 ng/µl of Arabidopsis A544 RNA was transcribed from A544 cDNA provided by Amersham using the Megascript protocol described above.
NIH20	180 ng/µl polyA RNA was - transcribed from DNA templates produced as PCR products from the individual NIH20 clones as described above. - the DNA templates were pooled and RNA was transcribed according to the Megascript protocol above. (~9 ng/µl of each clone is present in the pool assuming each template was equally transcribed)
SM	1 µg/µl of poly A+ RNA from human skeletal muscle cells, provided by Amersham.

RNA pooled as described below and reverse transcribed to produce single stranded DNA labeled with either Cy3 or Cy5; the RNA was degraded by RNAse H and One; and the DNA was purified using the GFX columns, according to the methods described above.

The RNA was pooled and labeled as follows:

Name	RNA Pooled	Label	Polynucleotide Concentration ng/label/slide
SM+NIH20 (red)	SM; NIH20; Arab.	Cy5-dCTP	SM = 987 NIH20 = 50 Arab. = 2.9
SM+NIH20 (green)	SM; NIH20; Arab.	Cy3-dCTP	SM = 987 NIH20 = 50 Arab. = 2.9
SM (green)	SM and Arab.	Cy3-dCTP	SM = 1000 Arab. = 2.9
RSM also referred to as "NIH20 (red)"	NIH20 and Arab.	Cy5-dCTp	NIH = 54 Arab. = 1.45

Mixtures comprising the following combinations of the RSM and samples were made as follows:

Mix 1 = SM+NIH20(red) and SM(green)

Mix 2 = SM+NIH20(green) and RSM

Mix 3 = SM(green) and RSM.

These mixtures were exposed to the target polynucleotides for hybridization as described above. The target polynucleotides included all twenty NIH20 clones and Arabidopsis A544.

The results of the assay are shown on the next page listing the ratios between each of the sample and the RSM and the ratio between the two samples as calculated from the RSM ratios and as determined empirically.

Example 2

Ratio Fidelity Validation

This experiment demonstrated that the ratios of the fluorescent signals produced by two mixtures correlates to the concentration ratio of polynucleotides of each mixture. Three mixtures were tested to determine the ratio of fluorescent signal produced in an Array format assay.

A DNA template comprising a T7 promoter was synthesized using the procedures described in Example 1 using Arabidopsis provided by Amersham. RNA was transcribed from the DNA template according to the procedure described in Example 1 using the Megascript kit. Further, labeled single

stranded DNA was reverse transcribed from the RNA according to Example 1 except, the RNA was degraded with RNAse H alone. Further, the hybridization to the target polynucleotide was performed using the procedure described in Example 1 except no DNA was added to block non-specific binding.

Three solutions were prepared with various concentrations of Cy3-dCTP labeled and Cy5-dCTP labeled DNA as shown below:

Experiment No.	Concentration of Cy3 Labeled cDNA	Concentration of Cy5 Labeled cDNA	Ratio of Concentration
29-6	250 copies per cell (cpc)	250 cpc	1:1
29-1	150 cpc	50 cpc	3:1
29-3	75 cpc	25 cpc	3:1

The DNA labeled as used for this experiment was A544 described above in Example 1. The single stranded DNA was transcribed as described in Example 1. Once transcribed, the DNA was purified from a GFX column as described above, but eluted in TE and the volume was not reduced.

Because the Cy5 labeled DNA produces a less intense signal than Cy3 labeled DNA, the Cy5 signals in all experiments were normalized by multiplying the Cy5 fluorescent volumes by 3.33. This normalization produced a 1:1 ratio between measurements of Cy3 and Cy5 fluorescent volumes when equal amounts of Cy3 and Cy5 labeled DNA were assayed. The 3.33 factor was determined empirically be assaying 35 samples that comprises 250 cpc of Cy3-labeled A544 and 250 copies of Cy5-labeled A544.

The results from the Array format assay are on the following page. The experiments show that the ratio of Cy3 to Cy5 measurements correlated with the ratio of polynucleotide concentration between the two mixtures, once the Cy5 measurements were normalized.

For example, the values for Cy3-labeled DNA in experiment number 29-1 varied from 64,469 to 280,557 fluorescent volumes, more than a 4 fold difference. However, the ratio of fluorescent measurements between the Cy3-labeled versus Cy5-labeled DNA varied by ±15.1%.

Further, the normalized ratio of experiment 29-1 averaged to 2.98 as compared to the 3:1 ratio of the polynucleotide concentrations between the Cy3- and Cy5-labeled DNA. Similarly, in experiment 29-3, the normalized ratio averaged to 2.91 as compared to the 3:1 ratio of the polynucleotide concentrations between the Cy3- and Cy5-labeled DNA.

Those skilled in the art will recognize, or be able to ascertain, using not more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such specific embodiments and equivalents are intended to be encompassed by the following claims.

All patents, published patent applications and publications cited herein are incorporated by reference as if set forth fully herein.

WHAT IS CLAIMED IS:

1. A method of comparing polynucleotide assay results comprising:

(a) providing a reference standard mixture (RSM);

(b) providing a first and second sample;

(c) providing a target polynucleotide;

(d) contacting the RSM and the first sample with the target polynucleotide under conditions that permit formation of RSM-target duplexes and sample-target duplexes by hybridization;

(e) determining the presence of duplexes;

(f) determining the ratio of duplexes formed between the first sample and the target nucleotide and the duplexes formed between the RSM and the target nucleotide;

(g) contacting the RSM and the second sample with the target polynucleotide under conditions that permit formation of RSM-target duplexes and sample-target duplexes by hybridization;

(h) determining the ratio of duplexes formed between the second sample and the target nucleotide and the duplexes formed between the RSM and the target nucleotide; and

(i) comparing the ratios of steps (f) and (h).

2. The method of claim 1, wherein the target polynucleotide is bound to a solid support.

3. The method of claim 2, wherein the target nucleotide is bound to a solid support in a high density Array format.

4. The method of claim 3, wherein the samples are contacted with a plurality of target polynucleotides.